PACE
Policy Analysis for California Education
CORE-PACE RESEARCH PARTNERSHIP

WORKING PAPER

A Middle School Drop:

Consistent Gender Differences in Students' Self-Efficacy

Erin M. Fahle
St. John's University

Monica G. Lee
Stanford University

Susanna Loeb
Brown University

Academic self-efficacy is a student's belief about their ability to learn or to perform within a school environment. This paper captures differential trends in academic self-efficacy by gender using self-efficacy survey data from five large districts in California from the 2014-15 through 2017-18 school years. We find that female students report significantly higher self-efficacy in elementary school compared to males. In middle school, students' self-efficacy declines for both genders; however, this drop is substantially greater for females, leading to significantly lower levels of reported self-efficacy for females than males from middle school onward. Despite large differences in average self-efficacy, this gendered pattern of drop-off occurs consistently across racial, socioeconomic, and academic subgroups. Average self-efficacy also varies significantly among schools; however, school demographics and culture and climate, as reported by students, are not strongly associated with the average female-male self-efficacy gap. Looking at how the general measure of academic self-efficacy corresponds with test scores, we find the drops in self-efficacy are most pronounced for low scoring students, and that changes in grade-to-grade test scores modestly correlate with changes in general academic self-efficacy.

VERSION: October 2019

Acknowledgements
This paper was produced as part of the CORE-PACE Research Partnership, which is focused on producing research that informs continuous improvement in the CORE districts (Fresno, Garden Grove, Long Beach, Los Angeles, Oakland, Sacramento City, San Francisco, and Santa Ana unified school districts) and policy and practice in California and beyond. This research was possible through the generous support of the S. D. Bechtel, Jr. Foundation and a grant from the Institute of Education Sciences (R305B090016) to the Board of Trustees of the Leland Stanford Junior University. We also thank the CORE Districts for partnering with us in this research, providing the data for this study, and giving feedback on earlier drafts of this study. Direct correspondence to Erin M. Fahle (*fahlee@stjohns.edu*). The viewpoints described herein are ours alone. PACE working papers are circulated for discussion and comment purposes and have not undergone the peer-review process that accompanies official PACE publications.

Introduction

Academic self-efficacy refers to students' beliefs about their ability to learn or to perform within a school environment (Bandura, 1986; Multon, Brown, & Lent, 1991). Students' self-efficacy is shaped by a combination of factors, including their personal successes or failures, the successes and failures of people they identify as similar to themselves, and their physical and emotional states (Bandura, 1982, 1997). Social cognitive theory postulates that an efficacious individual typically takes on a task with sustained effort, perceiving adversarial challenges as barriers that they have the ability to overcome. A less efficacious individual may have a weak sense of commitment or low aspirations about completing a challenging task, taking longer to bounce back from a setback or simply avoiding a task altogether.[1] Empirical evidence supports this theory, as many studies demonstrate that academic self-beliefs predict academic achievement outcomes (e.g., Multon et al., 1991; Bussey & Bandura, 1999; Pajares, 1996, 2005).[2] Disparities among groups in academic self-efficacy, therefore, may affect achievement gaps.

In this study, we focus on *gender disparities* in students' academic self-efficacy. Prior studies provide suggestive evidence of gender disparities in academic self-efficacy (e.g., Pajares, 2005), however, few have analyzed gender differences in broad academic attitudes, such as the belief that one has the ability to succeed in school. One notable exception is a study that we build on in this paper. West, Pier, Fricke, Hough, Loeb, Meyer, and Rice (2018b) show that although female students report significantly higher beliefs in their ability to do well in school (self-efficacy) in elementary grades compared to males, their efficacy drops more rapidly in middle school than that of males, leading to significantly lower levels of reported self-efficacy for females than males by the end of middle school and throughout high school.

The rapid middle school drop in self-efficacy and lower levels of self-efficacy throughout high school for females may have negative consequences for both their experiences in school and their later life outcomes. In this paper, we investigate which female students are experiencing these declines in self-efficacy and whether the changes are concentrated in some subgroups or school contexts. We draw on data for nearly 800,000 students across schools in

[1] Self-efficacy alone is not enough to result in positive performance; the individual must also hold the appropriate perception of the value of the task (as well as, the necessary skills). The expectancy-value theory framework contends that individuals who sense a high value or utility for a particular task will link their values to the appropriate outcome expectations to engage in behavior that they believe will lead to the desired outcome (Wigfield & Eccles, 1994; Wigfield, Tonks, & Eccles, 2004). Therefore, an efficacious person may have high expectancies, believing that they are capable of the type of tasks that will produce a desired outcome, and that the tasks ultimately bring them to a high-utility outcome.

[2] These associations vary, however, by how closely the measure of efficacy aligns with the measure of performance, and the type of performance measures, e.g., grades, standardized test scores, among others (Multon et al., 1991). "General" items, or items related to students' personal self-efficacy may not have the same predictive and explanatory power for academic performance as domain- or task-specific items (Bandura, 1997).

the California CORE districts to describe gender differences in academic self-efficacy, focusing on differences by demographic subgroups (income and race/ethnicity) and grade levels. We further explore whether gender disparities in self-efficacy are related to school culture and climate, and investigate their relationship with academic success. We provide evidence on female students' self-efficacy in order to inform programmatic efforts to boost self-efficacy, as well as to inform the use of these measures for school accountability.

Gender Gaps in Academic Self-Efficacy and Confidence

Most prior literature examines gender differences in domains that are stereotyped as male, such as math, science, and computer science. In these domains, students may hold gender-stereotypical beliefs about what they are 'supposed' to be good at, which could influence subsequent beliefs about their own academic success in such domains (Schunk & Pajares, 2002).

Indeed, researchers have consistently demonstrated the existence of gender disparities in self-beliefs about mathematics ability (Pajares, 2005). While there is some evidence that the gap in self-efficacy across these types of subjects can begin as early as in the first grade (Pajares & Miller, 1994; Schunk & Pajares, 2002), the majority of studies are on adolescents in middle and high school because the gender gap in both achievement and efficacy typically emerges and widens after elementary school (Huang, 2013; Hyde, Fennema, & Lamon, 1990). In a meta-analysis of 70 studies (126 unique data samples), Hyde, Ryan, Hopp, Fennema, and Frost (2006) find generally small, male-favoring gender differences in both self-confidence and attitudes in math. Moreover, while gaps existed among 11 to 14 year-olds, the differences in mathematics confidence was larger in older age groups (regardless of the scale used to measure confidence).

Analyses of the Trends in International Mathematics and Science Study (TIMSS) and the Programme for International Student Assessment (PISA) data also provide evidence that, among 14- to 16-year-old students, male students tend to have more positive math attitudes than girls even when they perform academically at the same level (Else-Quest, Hyde, & Linn, 2010). Research using the TIMSS data also demonstrates that in 8th grade, female students report lower self-efficacy in math, but that this difference does not fully account for the disparity in math performance (Louis & Mistele, 2012). Similarly, a recent study of the Philadelphia Adolescent Life Study data shows that male students have higher math self-concept in the 10th grade compared to females (Else-Quest, Mineo, & Higgins, 2013).

The evidence is less consistent when looking at other subject areas. Huang (2013), a meta-analysis of gender differences in self-efficacy, finds that male students reported higher self-efficacy in computer skills and social sciences compared to females, that there are no differences in science self-efficacy, and that female students have higher language arts self-efficacy. However, small samples of studies in the case of computer skills and social sciences limit the external validity of that particular finding.

In contrast to these subject-specific or task-specific measures, the current study analyzes a more domain-specific measure of students' self-efficacy, which captures their beliefs about their ability to succeed in school. Similar studies find that girls typically have higher efficacy in early grades and have mixed evidence about gender differences in middle school and beyond. Pajares and Graham (1999) and Pajares, Britner, and Valiante (2000) find that, compared to boys, girls reported greater average self-efficacy for self-regulation in elementary and middle school. Girls felt more confident in their ability to complete school activities such as homework and studying. Similarly, using an earlier release of the California CORE data used in the current study, West et al. (2018) find that female students are more confident in their ability to do well in school during elementary school. However, starting in middle school, the authors find males tend to report higher confidence in their ability to do well in school compared to females, and that this male-favoring self-efficacy gap continues through high school.

Academic Self-Efficacy, Student Demographics, and Context

Even less is known about how student demographics (race/ethnicity and socioeconomic status) or context may affect students' self-efficacy and gender differences in self-efficacy. Some studies find that white students tend to have higher self-efficacy than African American students (e.g., Pajares & Kranzler, 1995) while other studies indicate mixed findings on attitude differences by race (Kitsantas, Cheema, & Ware, 2011). There has been little exploration of whether gender differences in self-efficacy vary by racial group. One study by Saunders, Davis, Williams, and Williams (2004) finds female-favoring self-efficacy gaps in high school among a sample of African American students. Female students reported higher self-efficacy with respect to school completion and placed higher value on graduation relative to their male peers. This is suggestively in contrast to the findings of West et al. (2018) that finds male-favoring self-efficacy gaps in high school for all students; however, the use of different measures and samples makes it difficult to compare the results of the two studies.

With regard to context, Merolla (2017) shows that students in more disadvantaged communities have more variable self-efficacy relative to students in less disadvantaged contexts. Meelissen and Drent (2008) further provide evidence that school context may matter differentially by gender; however, their analysis focuses narrowly on computer science attitudes in the Netherlands. Their findings show that computer science attitudes of female students were positively related to the use of teacher-centered pedagogy (as compared to student-centered pedagogy) and to amount of computer experience of female teachers. The effects of each, however, were small.

Research Questions and Hypotheses

In this study, we ask: (1) How do self-efficacy and the self-efficacy gender gap change over grades?; (2) are these trends are consistent for students of different racial/ethnic and

socioeconomic status backgrounds?; (3) do they vary by school context?; and, (4) are they related to students' academic achievement?

Based on prior work by West et al. (2018), we hypothesize that the average level of self-efficacy among all students will be higher in elementary school and will drop off in middle school. Comparing males and females, we hypothesize faster declines in self-efficacy among female students relative to male students. We also hypothesize that this decline will be less steep for student subgroups who have had more opportunity for success – such as those who do not receive free or subsidized lunch in comparison to those eligible. Focusing on the intersection of gender with other subgroups, we further hypothesize that female-male gaps in self-efficacy may be more female-favoring for groups in which females have relatively higher academic performance than males such as for black, Hispanic, or economically disadvantaged students.

With regard to school context, we examine whether school demographics and school culture and climate (CC) are associated with different patterns of gender disparities in self-efficacy. We hypothesize that in schools with a more positive school climate, students' average self-efficacy will be higher and drop off less during middle school. We further expect that gender disparities in self-efficacy may also be smaller in these environments.

Finally, we hypothesize that prior academic performance will be positively related to one's efficacy levels given the importance of mastery experience in Bandura's theory of self-efficacy and prior research connecting the two (see Pajares, 1996 for a review). In the subsequent analysis, we focus on the association between grade-to-grade changes in self-efficacy and achievement to try to isolate a more direct, albeit not causal, connection between achievement and self-efficacy. Based on theory and prior literature, we hypothesize that grade-to-grade changes in self-efficacy will be positively correlated with grade-to-grade changes in academic performance.

Data

The data used in this paper come from the California CORE districts, which received a waiver granting exemption from constraints placed by the federal No Child Left Behind Act in order to build an innovative accountability system. This led to the implementation of a social-emotional learning (SEL) survey that systematically measures students' self-efficacy, self-management, growth mindset, and social awareness at scale to better assess efforts to support students.

We leverage data from five of the CORE districts. Our data include student-level social-emotional learning (SEL) survey responses, demographics, and academic achievement, as well as school-level culture and climate measures for students in grades three through twelve during the 2014–15 through 2017–18 school years. The student-level data also include identifying information about the school and district, which allows us to match the student-grade-year

observations to our set of time-invariant school characteristics. We supplement the CORE data with information from the Common Core of Data (CCD) that describes schools' demographics and enrollment.

Self-Efficacy

The SEL survey measures students' academic self-efficacy by using four items that ask students to rate how confident they feel about the following statements about school-related tasks on a 5-point scale from 1 (Not At All Confident) to 5 (Completely Confident): (a) I can earn an A in my classes; (b) I can do well on all my tests, even when they're difficult; (c) I can master the hardest topics in my classes; and, (d) I can meet all the learning goals my teachers set. We build a single composite of the four self-efficacy items by taking the average of the items for each student-year observation and standardizing to the overall mean and SD in 2015–16 SY.[3] In prior work, the self-efficacy scale was found to reliably distinguish between high and low efficacy students (Meyer, Wang, & Rice, 2018) and to be highly correlated with measures of self-esteem (West et al., 2018).

Demographic Characteristics

We use time-invariant variables indicating student gender, race/ethnicity, and socioeconomic disadvantage. Gender classifies students as female or male. Race/ethnicity is captured in five categories: non-Hispanic white, black, Hispanic, Asian (contains Asian, Filipino, and Pacific Islander),[4] or other race/ethnicity (contains Native American, multi-race, and not specified race/ethnicity).[5] Socioeconomic disadvantage is equivalent to ever being eligible to receive free or reduced-price lunch (FRPL) during the years in our sample.

Culture and Climate Survey Measures

We use culture and climate (CC) survey response data from elementary and secondary school students in the 2015–16 school year.[6] While elementary and high school students received different versions of the survey, both versions measured the same four constructs: (1) Climate of Support for Academic Learning (6 questions); (2) Sense of Belonging – School Connectedness (5 questions); (3) Knowledge and Fairness of Discipline, Rules and Norms

[3] We select this reference year because we observe a large and representative sample of survey responses from all participating districts.

[4] Due to sample size constraints, we group students identified as Asian, Filipino, and Pacific Islander into a single 'Asian' category regardless of the practices in the individual districts.

[5] Students who reported both black and Hispanic, the student was coded as black.

[6] Although the survey was administered annually during out sample time frame, we use only 2015-16 CC survey data for two reasons. First, CC survey responses are missing from 1 to 2 districts in 2016-17 and 2017-18 SYs, and there are considerably fewer responses from 2014-15, which was a pilot year. Second, based on mean survey responses by year, we observe that responses are fairly stable across years.

(9 questions); and (4) Sense of Safety (7 questions). The questions that measure the first three constructs are largely identical. However, the survey items intended to measure the Sense of Safety construct vary significantly across grades due to differences in the perceived potentially dangerous instances younger vs. older students may encounter or report. Therefore, we do not use this construct in our analyses.

We create three school-level culture and climate composites – Academics, Belonging, and Discipline, for short. We first standardize the responses to each item to have a mean of zero and a standard deviation of one and calculate student-level averages across the items.[7] Then, we average the individual responses within schools and re-standardize at the school-level. In order to mitigate extreme values, we only use school-averages where at least 100 students responded to the surveys. The pairwise correlations between the three measures are all approximately .88.

School Demographic Characteristics

Using the CCD, we calculate the average proportion of each racial/ethnic group and the average proportion of students receiving FRPL served by each school in our sample during the four-year period we analyze. We also create a measure of the average grade-level enrollment for the school (averaged over grades and years).

Analytic Sample

We restrict our primary analytic sample to include students who have complete student-level SEL and demographic data, as well as school-level CC and demographic data.[8] The final sample consists of 1,343,715 student-year observations from 796,581 students in 813 schools across four school years (2014–15 through 2017–18). As a robustness check, we also run our models using the sample of students who were observed at least two times in our panel data (shown in Table A1). Table 1 provides the average characteristics of students (Panel 1) and schools (Panel 2) in the sample. The students are approximately half female and half male, and overwhelmingly Hispanic (72%) and low-income (84%).

[7] We standardize items prior to aggregating because the number of Likert response options varies across items.

[8] We removed students who were observed multiple times within a single year. We also removed students who did not make "normal" grade progress (i.e. were retained or skipped grades). In total, these restrictions removed 1.7% of the total observations (23,491 out of 1,367,206).

Table 1. Summary of Student and School Covariate Data

	Full Sample	
	Mean	SD
Panel 1. Student-level Summary Statistics		
Female	0.49	(0.50)
Male	0.51	(0.50)
White	0.08	(0.26)
Black	0.08	(0.28)
Hispanic	0.72	(0.45)
Asian	0.10	(0.30)
Other	0.02	(0.14)
FRPL	0.84	(0.37)
Self-Efficacy Composite	3.47	(1.02)
Panel 2. School-level Summary Statistics		
White (%)	0.08	(0.12)
Black (%)	0.08	(0.11)
Hispanic (%)	0.72	(0.25)
Asian (%)	0.09	(0.14)
Other (%)	0.02	(0.03)
FRPL (%)	0.79	(0.19)
Average Enrollment per Grade (100s)	1.75	(1.57)
Culture & Climate Survey Constructs		
Belonging Composite	3.53	(0.22)
Learning Composite	3.69	(0.27)
Discipline Composite	3.54	(0.24)

Note: Panel 1 shows descriptive statistics for the 796,581 students in the sample. Panel 2 shows descriptive statistics for 813 schools. The self-efficacy composite is an average of four items that pertain to this construct. The culture and climate survey constructs are composites that measure school averages of standardized items pertaining to each construct.

Test Score Sample

We construct a second sample that restricts to observations with: (1) prior and current year self-efficacy scores and (2) prior and current-year Smarter Balanced Assessment Consortium (SBAC) test scale scores in math. These data allow us to look at differences in self-efficacy trends by achievement level, categorized here as “high scoring,” “mid scoring,” or “low scoring,” as well as to understand the relationship between concurrent changes in test scores and changes in self-efficacy. We categorize a student in the current year and grade as “high scoring” (“low scoring”) if they scored above the 75th (below the 25th) percentile in the within-grade standardized score distribution in the prior grade. A student is categorized as “mid scoring” if they are in the middle half of the distribution in the prior grade and year (between

the 25th and 75th percentiles). Because SBAC grade-level testing only occurs consecutively in 3rd through 8th grade, observations in this sample are only for grades 4 through 8 in the 2015-16 through 2017–18 SYs. The sample is similar demographically to the main analytic sample. Descriptive characteristics can be found in Appendix Table A4.

Methods

We use hierarchical linear modeling in order to estimate (1) gender differences in students' self-efficacy and in the change in students' self-efficacy across grades; and, (2) differential levels and trends across student and school groups. This approach allows us to estimate the accuracy of our coefficients and thus test whether the groups are statistically discernable.

Our primary specification is a three-level hierarchical linear model where student-grade-year observations are nested in students and students are nested in schools. We estimate fixed non-parametric grade trends for the average male self-efficacy ($\beta_{200}, \beta_{300}, \dots, \beta_{1000}$) and the average female-male self-efficacy gap ($\beta_{210}, \beta_{310}, \dots, \beta_{1010}$), allowing them to vary freely by grade level. We include one student-level random effect on the intercept, average self-efficacy (r_{0ij}). We also include four school-level random effects on the intercept (u_{00j}), the female-male gap in self-efficacy (u_{02j}), the linear self-efficacy grade trend (u_{10j}), and the linear female-male self-efficacy gap grade trend (u_{10j}) to capture variability across schools.

$$sel_{ijy} = \alpha_{0ij} + \alpha_{1ij} grade_{ijy} + \alpha_{2ij} g3 + \alpha_{3ij} g4 + \cdots + \alpha_{10ij} g12 + e_{ijy}$$

$$\begin{aligned}
\alpha_{0ij} &= \beta_{00j} + \beta_{010} cohort_{ij} + \beta_{02j} female_{ij} + r_{0ij} \\
\alpha_{1ij} &= \beta_{10j} + \beta_{11j} female_{ij} \\
\alpha_{2ij} &= \beta_{200} + \beta_{210} female_{ij} \\
&\vdots \\
\alpha_{10ij} &= \beta_{1000} + \beta_{1010} female_{ij}
\end{aligned}$$

$$\begin{aligned}
\gamma_{00j} &= \boldsymbol{S_j}\boldsymbol{\Gamma_{000}} + u_{00j} \\
\gamma_{02j} &= \boldsymbol{S_j}\boldsymbol{\Gamma_{020}} + u_{02j} \\
\gamma_{10j} &= \boldsymbol{S_j}\boldsymbol{\Gamma_{100}} + u_{10j} \\
\gamma_{11j} &= \boldsymbol{S_j}\boldsymbol{\Gamma_{110}} + u_{11j}
\end{aligned}$$

$$e \sim N(0, \sigma^2), r_{0ij} \sim N(0, \tau_1^2), \quad \mathbf{U} \sim MVN\left(0, \boldsymbol{\tau}_\mathbf{2}^\mathbf{2}\right) \tag{1}$$

In the above equation, sel_{ijy} is the composite self-efficacy measure for student i, in school j, in year y; $grade_{ijy}$ is a continuous grade variable centered at 7.5; $g3, g4, \ldots, g12$ are a set of grade dummies (representing grades 3 through 12); $cohort_{ij}$ is a continuous variable equal to year minus grade and centered at 2008 that indicates the year a student completed Kindergarten; $female_{ij}$ is an indicator equal to one if the student is female; and, $\boldsymbol{S_j}$ is a vector of school characteristics.

We estimate this specification first for all students without school covariates (our baseline model) and then separately for each student subgroup (e.g., white students, black students, Hispanic students, Asian students, low-income students, and non-low-income students). For comparability across the samples, we do not alter the structure of the model regardless of the statistical significance of the random effects.[9] These models provide an estimate of (1) whether there are gender differences in the levels and trends of students' self-efficacy over grades; (2) whether these levels and trends vary among key subgroups of students; and, (3) whether the levels and trends vary among schools.

To each of these models we add a set of the CC survey composites (as well as a set of school demographic control variables) in $\boldsymbol{S_j}$ to explore whether they explain the variance among schools in self-efficacy gaps. We estimate the proportion of explained variance relative to the corresponding baseline model with no school-level covariates.

Test-Score Analyses

To explore whether prior achievement is related to trends in self-efficacy we estimate a variant of Equation (1) separately for the low-, mid-, and high-scoring subgroups of students with student observations nested in school-by-year clusters nested in schools. We further explore whether standardized grade-to-grade changes in test scores are related to standardized changes in self-efficacy using a fixed effects model, shown below, and estimated separately for all grade pairs (e.g., for 4th to 5th grade; for 5th to 6th grade; etc.):

$$scorechange_i^g = \beta_1 selchange_i^g + \beta_2 female_i + \beta_3 selchange_i^g * female_i + \boldsymbol{\Lambda}_y + \boldsymbol{\Gamma}_s + e_i \quad (2)$$

In this equation, $scorechange_i^g$ is the standardized difference in a students' SBAC math scores between consecutive grades ($g-1, g; g \in [4,8]$) for student i; $selchange_i^g$ is the

[9] In the model for black students only, the random effects u_{02j} and u_{11j} are not significant at the 0.05 level. In the model for white students only, the random effect u_{02j} is not significant at the 0.05 level. in the models for high, mid, and low scoring students, the random effects on $r_{1j}$and r_{3j} are not significant (likely due to the small sample size) but have been retained for comparability. Random effects in all other models (without covariates) are significant.

standardized difference in a students' self-efficacy between consecutive grades; $female_i$ is an indicator that student i is female; and, $\boldsymbol{\Lambda}_y$ and $\boldsymbol{\Gamma}_s$ are year and school fixed effects, respectively.[10]

Of interest in this model are the coefficients β_1, which provides an estimate of the relationship (if any) between grade-to-grade changes in self-efficacy and test scores, and β_3, which identifies whether this relationship differs for females and to what extent.

Results

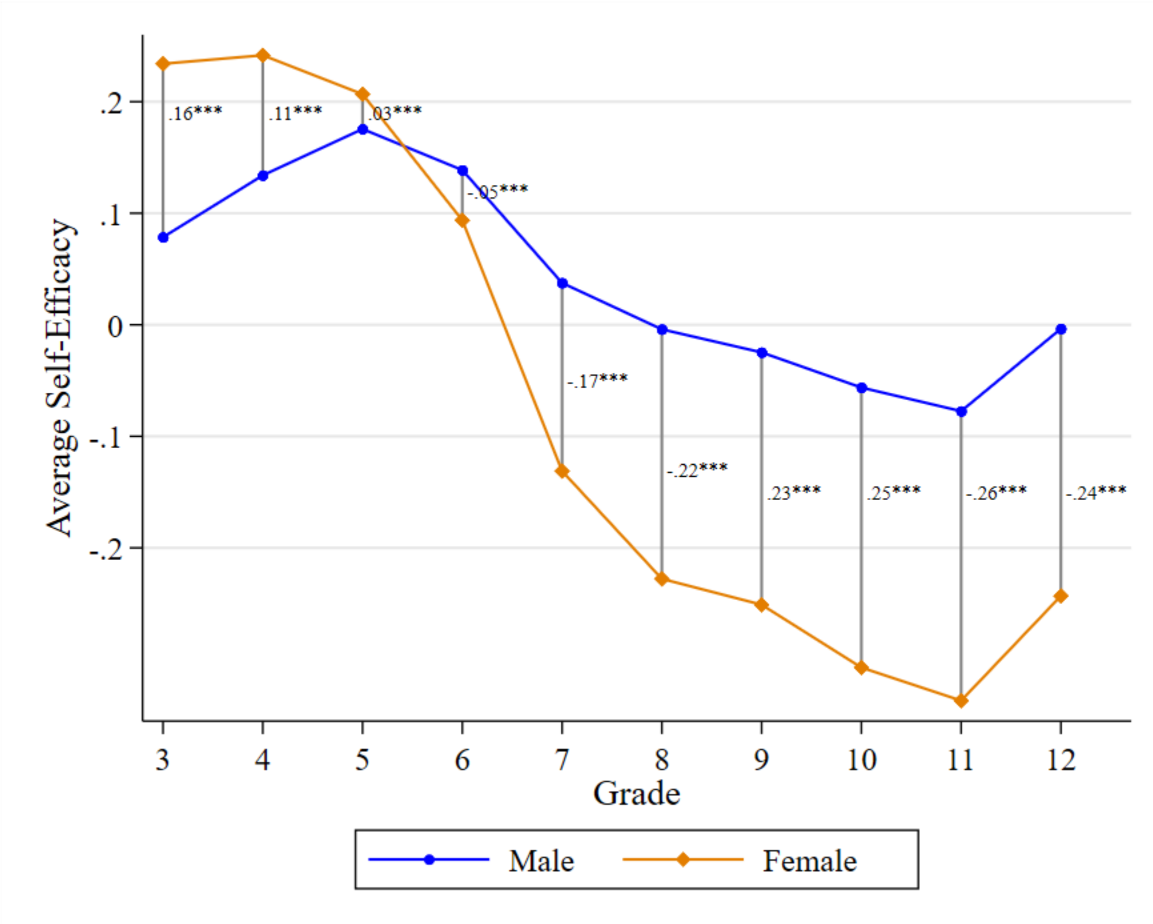
Gender Differences in Self-Efficacy

We report changes in the self-efficacy composite in standard deviation units. We standardized the composite across grades to the 2015–16 SY, such that zero represents the average self-efficacy of students across all grades in that year. We chose to not standardize within grade, so that the decline in absolute self-efficacy across grades is evident. Male students are the comparison group. Self-efficacy for male students is about .079 in third grade and increases significantly through fifth grade to .176 SD; (see Figure 1, Table A1). However, at the transition into middle school self-efficacy quickly declines, such that by eighth grade male students report self-efficacy equal to the average (for all students across grades). Male self-efficacy remains at or below average throughout high school. The steepest decline, from .139 SD to .038 SD, occurs from grades six to seven, and the second steepest decline, from .038 SD to -.004 SD, follows immediately from grades seven to eight.

[10] We calculate standardized grade-to-grade differences in students' self-efficacy and test scores by calculating the grade-to-grade difference in either test scores or self-efficacy scores in the metric of each assessment. Then we standardize the differences within the sample.

$$scorechange_i^g = std\left(score_{i,g} - score_{i,g-1}\right)$$
$$selchange_i^g = std\left(sel_{i,g} - sel_{i,g-1}\right)$$

Figure 1. Changes in Female-Male Self-Efficacy Gaps Over Grades



Notes: The units are standardized self-efficacy composite score points. We report the estimated coefficients on the interactions between the grade fixed effects and the female indicator. Observations are from third through twelfth grade students across five districts in the 2014–2015 and 2017–18 school years. Star(s) indicate a statistically significant gap between males and females. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

In all grades, there is a statistically significant difference in the average self-efficacy of female students relative to male students. Compared to their male counterparts, female students report higher self-efficacy in elementary school and lower self-efficacy in high school (Table A1, bottom panel). In elementary grade levels, female students' self-efficacy is up to .155 SD higher than males (in grade 3), but it drops off more rapidly than male self-efficacy during middle school yielding a large male-favoring self-efficacy gap of -.22 SD in eighth grade. This gap remains stable for the remainder of high school.

Variability Across Demographic Subgroups

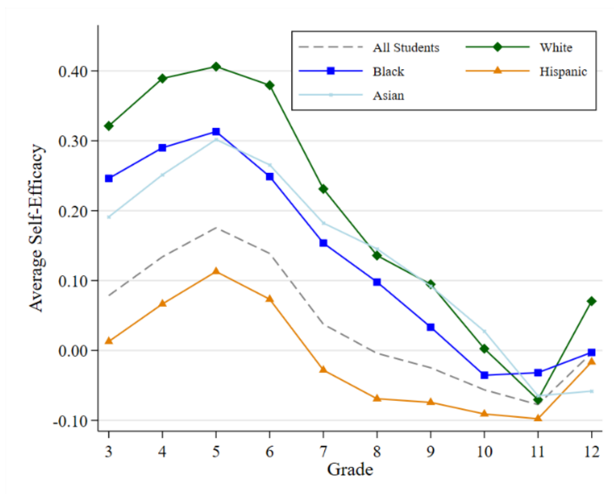
Among the racial/ethnic subgroups in our sample, white students on average report the highest average efficacy (.321 SD), followed by black students (.246 SD), Asian students

(.191 SD) and Hispanic students (.013 SD) in grade 3 (Figure 2, Panel A; Table A2). Differences between racial/ethnic subgroups get smaller over grades, becoming nearly identical by grade 11. Students who are not low-income report significantly higher average efficacy than eligible students in grade 3 (.150 SD vs .046 SD); again, these differences persist until about grade 11 where the efficacy between the two groups is indistinguishable (Figure 2, Panel C). Similar to the results for all students, students of each demographic subgroups experience sharp declines in efficacy beginning in 6th grade that, for the most part, begin to level off in high school.

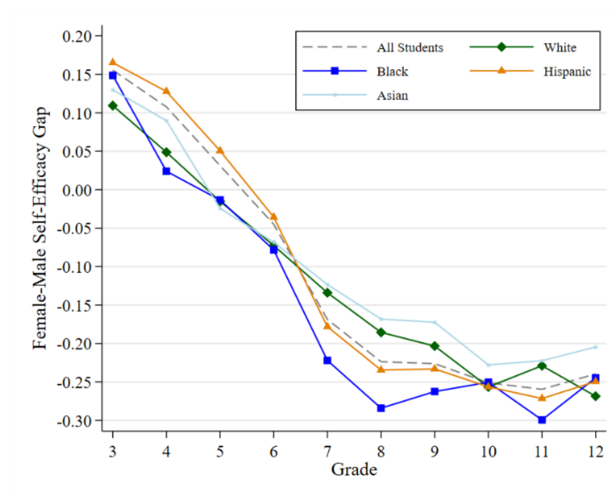
Regardless of student characteristics, female students report higher self-efficacy than male students in third grade (positive gap), there is a steeper drop in self-efficacy for females compared to males, and male students report higher average self-efficacy than female students (negative gap) starting in 6th grade and continuing through high school (Figure 2, Panels B and D; Table A2).

Figure 2. Changes in Average Male Self-Efficacy and Female-Male Self-Efficacy Gaps Over Grades by Demographic Subgroup

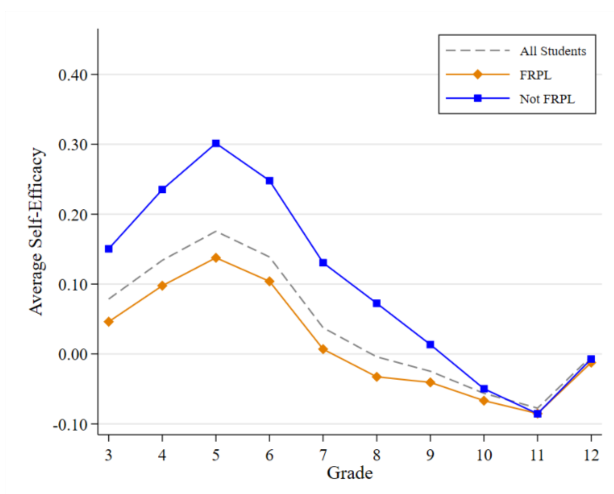
A.



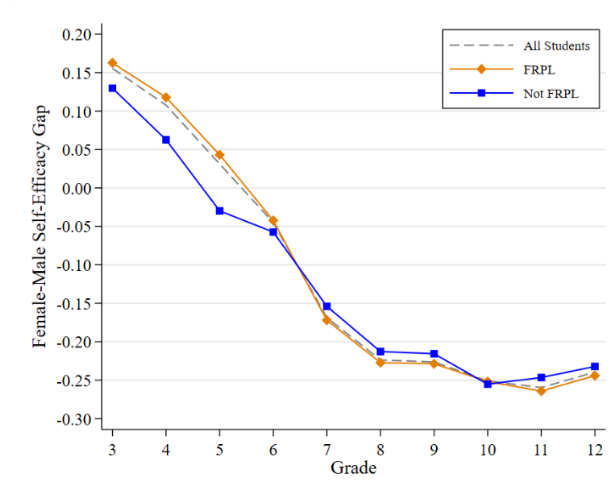
B.



C.



D.



Notes: The units are standardized self-efficacy composite score points. The points reported here are the estimated coefficients on the interactions between the grade fixed effects and the female indicator. Observations for the race/ethnicity and FRPL status subgroups are from third through twelfth grade students across five districts between the 2014–15 and 2017–18 school years.

Variability Across Schools

We estimate significant variance in the average self-efficacy between schools (SD = .20) and between students (SD = .63); approximately 9% of the total variance in self-efficacy is between-schools (full results in Table A1).[11] In other words, there are meaningful differences between schools in students' average self-efficacy and school-level factors may contribute to students' average self-efficacy. In contrast, there is little true between-school variance in the female-male efficacy gap (SD = .05), as well as little to no between-school variance in the average and gap grade trends (SD =.05 and SD =.01, respectively). Given we estimate little variance between schools in the female-male gap and gap trend, we do not focus on these parameters in this discussion.

School culture and climate explains little of the variance in average efficacy between schools (4% to 9%, Table 2). We find average self-efficacy is higher in schools with more supportive academic learning climates (.051, $p < .001$), in schools where students report a higher sense of belonging (.059, $p < .001$) and in schools where students perceive discipline is fair (.035, $p < .001$). These associations of self-efficacy with supportive learning climates and perceptions of fair discipline are slightly more pronounced for female students relative to male students. School demographics (the percentage of students by race/ethnicity and FRPL), however, explain approximately 20% of the between-school variance in average self-efficacy and the self-efficacy trend (Table A4). Therefore, racial/ethnic and economic student sorting between schools may partially explain the variability in school average efficacy and trends, while the role of school culture and climate—in so much as we are able to measure it—appears limited.

[11] Potential measurement error in the self-efficacy composite, especially given that it is self-reported, may increase the appearance of variation between students within the same school and reduce the percentage of variance across schools.

Table 2. Relationship of Self-Efficacy to School Culture and Climate

	Composite					
	Learning		Belonging		Discipline	
Average Male Self-Efficacy						
Composite	0.051	***	0.059	***	0.035	***
	(0.008)		(0.008)		(0.008)	
Composite-x-Grade Slope	-0.002		-0.004	+	-0.002	
	(0.002)		(0.002)		(0.002)	
Female-Male Self-Efficacy Gap						
Composite	0.009	**	0.002		0.005	+
	(0.003)		(0.003)		(0.003)	
Composite-x-Grade Slope	0.000		0.000		-0.003	*
	(0.001)		(0.001)		(0.001)	
Between-School Intercept R^2	0.058		0.086		0.036	
Between-School Grade Slope R^2	0.034		0.000		0.000	
Between-School Female Slope R^2	0.001		0.017		0.009	
Between-School Female-x-Grade Slope R^2	0.062		0.014		0.010	

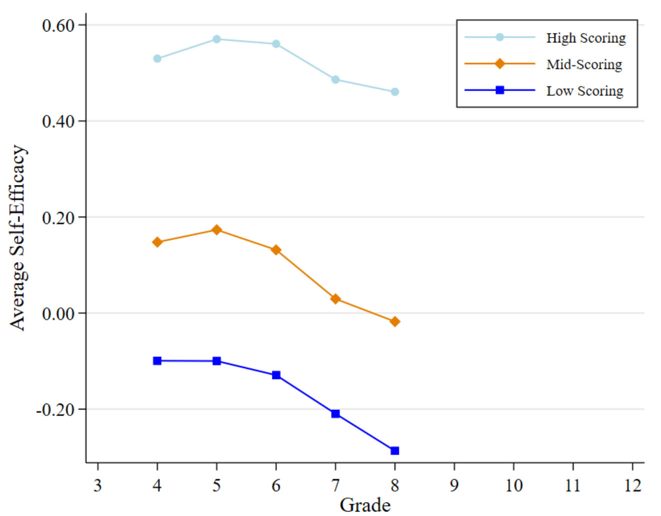
Note: Individual models were fitted for each composite by student subgroup. Terms included but not shown in the table are a cohort variable centered at 2015, a nonparametric grade trend, and a grade variable centered at 7.5. Model results reported portray trends in self-efficacy based on school-level composites of 2015–2016 SY survey responses from students. Survey construct variables are standardized and averaged within schools. In all analyses there are 1,343,715 observations, 796,581 students, and 813 schools. Standard errors in parentheses; + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Self-Efficacy and Test Scores

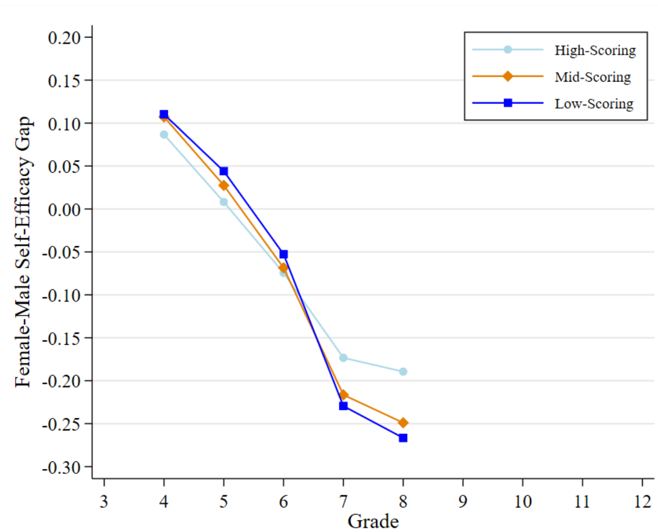
High scoring students report the highest average efficacy, followed by mid-scoring students, and low scoring students in all grades and years. Among high scoring students, self-efficacy declines in middle school, but remains well above zero (the average self-efficacy for grade 3–12 students in the 2015–16 SY) at approximately .46 SD. In contrast, low scoring students' self-efficacy begins slightly below average and declines rapidly to approximately -.3 SD. Despite these stark differences in average levels of efficacy, the gap in self-efficacy is very similar across the three groups (Figure 3). Gaps are female-favoring in grade 4, near zero in grade 5 and rapidly increase in favor of males to -.2 to -.3 SD by the end of middle school. In grades 7 and 8, the gaps in favor of males for low scoring and mid-scoring students are about .05 SD larger than the gap favoring males among high scoring students.

Figure 3. Changes in Average Male Self-Efficacy and Female-Male Self-Efficacy Gaps Over Grades by Prior Achievement

A.



B.



We further find there is a small, positive association between the change in efficacy and the change in test scores in all grades (Table A5). In other words, a student who experienced a 1 SD larger, positive change in efficacy also tends to also have a larger, positive change in test scores on the order of .015 to .062 SD. These coefficients are small in magnitude, and do not suggest a strong connection between efficacy and math test scores. Moreover, there is no evidence that the association differs for female students. Notably, our self-efficacy measure is not specific to mathematics. This lack of specificity likely limits its predictive power in explaining changes in standardized math test scores. A better measure may be students course grades, as these outcomes are more directly assessed in the efficacy items.

Discussion

This paper describes variation in students' academic self-efficacy—their beliefs in their ability to perform within a school environment. Prior research shows that students experience a drop in academic self-efficacy during middle school that is particularly strong for female students, resulting in lower self-efficacy for girls than boys from middle school through high school. In this paper we provide evidence on the consistency of this pattern across student groups as defined by demographics, achievement level, and school of attendance in order to better understand whether a subset of groups is driving this overall trend.

Across all groups that we assess, we find unusual consistency in the differential drop in academic self-efficacy during middle school. On average, students who are non-white, low-achieving, and/or poor demonstrate somewhat greater drops than their peers, and students who show an improvement in their academic achievement tend to experience smaller drops. Nevertheless, the differential nature of the self-efficacy drop in males and females is essentially universal across all student groups. Similarly, while schools vary meaningfully in their students' average level of self-efficacy, they do not differ nearly as much in self-efficacy gender gaps or trends.

An unexpected lack of variation across subgroups and school context does not imply that this gender gap is unalterable, nor that school practices could not reduce it. On average, male and female students have similar access to educational resources within their homes and schools. So, it is unlikely that the self-efficacy gap is related to systematic differences in schooling or in male and female students' access to experiences or institutions. Moreover, our data show large male-female self-efficacy gaps among similarly performing male and female students (e.g., high-scoring, mid-scoring, and low-scoring). So, meaningful gaps in prior performance are unlikely to explain the disparity.

Instead, what occurs *within* male and female students' shared academic experiences likely affects their self-efficacy in different ways. For example, while female and male students are often in the same classrooms, they may be treated differently inside of those classrooms (Jones & Dindia, 2004). Moreover, males and females tend to evaluate themselves differently given similar feedback (Pomerantz, Altermatt, & Saxon, 2002). Research also provides some

evidence that female students' self-efficacy beliefs are particularly responsive to social persuasion—how others perceive and describe their abilities (Zeldin & Pajares, 2000). This finding suggests that the messaging female students receive, or how they interpret the messages they receive about their abilities, is critical for the development of their self-beliefs; and, that ensuring female students receive positive messaging about their performance in classroom environments may be a first step to closing these persistent gaps. Future research should continue to identify factors that influence self-efficacy, so that school-based interventions – particularly in classrooms across middle schools, where the self-efficacy gap begins – can be implemented and tested.

Although not the focus of this paper, we find meaningful differences in the average levels of self-efficacy between minority and white students, between poor and non-poor students, and between low- and high-achieving students. In all cases, groups that are more disenfranchised within the learning environment appear to be more susceptible to self-doubt and lack of confidence. These results fit with prior research which suggests that white students tend to have higher self-efficacy than students of color (Pajares & Kranzler, 1995) and that low-SES students tend to have lower self-esteem than their high-SES peers (e.g., Bolger, Patterson, Thompson, & Kupersmidt, 1995). One possible explanation for this pattern is that marginalized students have fewer high-quality opportunities to learn and, therefore, to build their academic confidence relative to their peers. Interventions that focus on reducing opportunity gaps could be effective at reducing disparities in self-efficacy among these subgroups.

The fact that the data used in this paper were derived from an effort to incorporate SEL measures into accountability merits a final comment. Our work highlights potential negative implications for using self-efficacy or other SEL measures for higher-level accountability. A non-trivial proportion of the variation in self-efficacy across school contexts is explained by student background, and male-female efficacy gaps vary little across schools. Without a better understanding of the policies and practices that schools can adopt to bolster academic confidence, it might be counterproductive to hold schools responsible for their students' self-efficacy (Usher & Pajares, 2008; Marsh et al., 2018). Evidence provided in this paper points to the need for further and more careful examination of links between students' SEL skills, interventions used at varying levels to improve them, and the role our classrooms and schools play in SEL development.

References

Bandura, A. (1982). The assessment and predictive generality of self-percepts of efficacy. *Journal of Behavior Therapy and Experimental Psychiatry*. https://doi.org/10.1016/0005-7916(82)90004-0

Bandura, A. (1997). *Self-efficacy: The exercise of control.* New York, NY, US: W H Freeman/Times Books/ Henry Holt & Co.

Bolger, K. E., Patterson, C. J., Thompson, W. W., & Kupersmidt, J. B. (1995). Psychosocial adjustment among children experiencing persistent and intermittent family economic hardship. *Child Development*. https://doi.org/10.1111/j.1467-8624.1995.tb00926.x

Bussey, K., & Bandura, A. (1999). Social cognitive theory of gender development and differentiation. *Psychological Review*, *106*(4), 676–713. https://doi.org/10.1037/0033-295x.106.4.676

Else-Quest, N. M., Hyde, J. S., & Linn, M. C. (2010). Cross-national patterns of gender differences in mathematics: A meta-analysis. *Psychological Bulletin*, *136*(1), 103–127. https://doi.org/10.1037/a0018053

Else-Quest, N. M., Mineo, C. C., & Higgins, A. (2013). Math and science attitudes and achievement at the intersection of gender and ethnicity. *Psychology of Women Quarterly*, *37*(3), 293–309. https://doi.org/10.1177/0361684313480694

Huang, C. (2013). Gender differences in academic self-efficacy: A meta-analysis. *European Journal of Psychology of Education*, *28*(1), 1–35. https://doi.org/10.1007/s10212-011-0097-y

Hyde, J. S., Fennema, E., & Lamon, S. J. (1990). Gender differences in mathematics performance: A meta-analysis. *Psychological Bulletin*, *107*(2). https://doi.org/10.1037/0033-2909.107.2.139

Hyde, J. S., Ryan, M., Hopp, C., Fennema, E., & Frost, L. A. (2006). Gender comparisons of mathematics attitudes and affect: A meta-analysis. *Psychology of Women Quarterly*, *14*(3), 299–324. https://doi.org/10.1111/j.1471-6402.1990.tb00022.x

Jones, S. M., & Dindia, K. (2004). A meta-analytic perspective on sex equity in the classroom. *Review of Educational Research*, *74*(4), 443–471. https://doi.org/10.3102/00346543074004443

Kitsantas, A., Cheema, J., & Ware, H. W. (2011). Mathematics achievement: The role of homework and self-efficacy beliefs. *Journal of Advanced Academics*, *22*(2), 310–339. https://doi.org/10.1177/1932202X1102200206

Louis, R. A., & Mistele, J. M. (2012). The differences in scores and self-efficacy by student. *International Journal of Science and Mathematics Education*, (November 2011), 1163-1190.

Meelissen, M. R. M., & Drent, M. (2008). Gender differences in computer attitudes: Does the school matter? *Computers in Human Behavior*. https://doi.org/10.1016/j.chb.2007.03.001

Merolla, D. M. (2017). Self-efficacy and academic achievement. *Sociological Perspectives*, *60*(2), 378–393. https://doi.org/10.1177/0731121416629993

Meyer, R. H., Wang, C., & Rice, A. B. (2018). *Measuring students' social-emotional learning among California's CORE districts: An IRT modeling approach.* Policy Analysis for California Education. Retrieved from http://www.edpolicyinca.org/sites/default/files/Measuring_SEL_May-2018.pdf

Multon, K. D., Brown, S. D., & Lent, R. W. (1991). Relation of self-efficacy beliefs to academic outcomes: A meta-analytic investigation. *Journal of Counseling Psychology*, *38*(1), 30-38. https://doi.org/10.1037/0022-0167.38.1.30

Pajares, F. (2005). Gender differences in mathematics: self-efficacy beliefs. In *Gender differences in mathematics: An integrative psychological approach* (pp. 294–315). https://doi.org/doi:10.1017/CBO9780511614446.013

Pajares, F., Britner, S. L., & Valiante, G. (2000). Relation between achievement goals and self-beliefs of middle school students in writing and science. *Contemporary Educational Psychology*. https://doi.org/10.1006/ceps.1999.1027

Pajares, F., & Graham, L. (1999). Self-efficacy, motivation constructs, and mathematics performance of entering middle school students. *Contemporary Educational Psychology*, *24*, 124–139. https://doi.org/10.1006/ceps.1998.0991

Pajares, F., & Kranzler, J. (1995). Self-efficacy beliefs and general mental ability in mathematics. *Contemporary Educational Psychology*, *20*, 426–443.

Pajares, F., & Miller, M. D. (1994). Role of self-efficacy and self-concept beliefs in mathematical problem solving: A path analysis. *Journal of Educational Psychology*, *86*(2), 193–203. https://doi.org/10.1037/0022-0663.86.2.193

Pomerantz, E. M., Altermatt, E. R., & Saxon, J. L. (2002). Making the grade but feeling distressed: Gender differences in academic performance and internal distress. *Journal of Educational Psychology*, *94*(2), 396–404. https://doi.org/10.1037/0022-0663.94.2.396

Saunders, J., Davis, L., Williams, T., & Williams, J. H. (2004). Gender differences in self-perceptions and academic outcomes: A study of African American high school students. *Journal of Youth and Adolescence*, *33*(1), 81–90. https://doi.org/10.1023/A:1027390531768

Schunk, D. H., & Pajares, F. (2002). The development of academic self-efficacy. *Development of Achievement Motivation*, *1446*, 15–31. https://doi.org/10.1016/B978-012750053-9/50003-6

West, M. R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R. H., & Rice, A. B. (2018a). *Trends in student learning: Evidence from the CORE districts.*

West, M. R., Pier, L., Fricke, H., Hough, H., Loeb, S., Meyer, R. H., & Rice, A. B. (2018b). Trends in student learning: Evidence from the CORE districts. *PACE Working Paper*.

Wigfield, A., & Eccles, J. S. (1994). Children's competence beliefs, achievement values, and general self-esteem: Change across elementary and middle school. *The Journal of Early Adolescence*. Journal of Early Adolescence. https://doi.org/10.1177/027243169401400203

Wigfield, A., Tonks, S., & Eccles, J. S. (2004). Expectancy value theory in cross-cultural perspective. In D. M. McInerney & S. Van Etten (Eds.), *Big Theories Revisited* (pp. 165-198). Greenwich, CT: Information Age Publishing.

Zeldin, A. L., & Pajares, F. (2000). Against the odds: Self-efficacy beliefs of women in mathematical, scientific, and technological careers. *American Educational Research Journal*, *37*(1), 215–246.

Appendix A

Table A1. Baseline Model Results by Student Sample

	All Students		Students Observed in Multiple Years	
Average Male Self-Efficacy				
Grade 3	0.079	***	0.097	***
	(0.016)		(0.018)	
Grade 4	0.134	***	0.154	***
	(0.013)		(0.015)	
Grade 5	0.176	***	0.209	***
	(0.012)		(0.014)	
Grade 6	0.139	***	0.179	***
	(0.012)		(0.013)	
Grade 7	0.038	**	0.086	***
	(0.012)		(0.013)	
Grade 8	-0.004		0.051	***
	(0.012)		(0.013)	
Grade 9	-0.025		0.051	**
	(0.018)		(0.020)	
Grade 10	-0.056	**	-0.009	
	(0.019)		(0.021)	
Grade 11	-0.077	***	-0.050	*
	(0.022)		(0.024)	
Grade 12	-0.004		0.009	
	(0.025)		(0.027)	
Female-Male Self-Efficacy Gap				
Grade 3	0.155	***	0.156	***
	(0.007)		(0.008)	
Grade 4	0.108	***	0.110	***
	(0.005)		(0.006)	
Grade 5	0.031	***	0.028	***
	(0.006)		(0.007)	
Grade 6	-0.045	***	-0.037	***
	(0.006)		(0.007)	
Grade 7	-0.169	***	-0.170	***
	(0.006)		(0.007)	
Grade 8	-0.224	***	-0.232	***
	(0.006)		(0.008)	
Grade 9	-0.226	***	-0.229	***
	(0.007)		(0.010)	
Grade 10	-0.251	***	-0.265	***
	(0.008)		(0.009)	
Grade 11	-0.260	***	-0.274	***
	(0.009)		(0.010)	
Grade 12	-0.240	***	-0.255	***
	(0.009)		(0.011)	
Number of Observations	1343715		945145	
Number of Students	796581		398011	
Number of Schools	813		813	
Between-Student Intercept Variance	0.400		0.394	
Between-School Intercept Variance	0.040		0.041	
Between-School Grade Slope Variance	0.002		0.003	
Between-School Female Slope Variance	0.003		0.003	
Between-School Female-x-Grade Slope Variance	0.000		0.000	
Reliability of Student Intercept	0.527		0.628	
Reliability of School Intercept	0.684		0.618	
Reliability of School Female Slope	0.164		0.148	
Reliability of School Grade Slope	0.598		0.548	
Reliability of School Female-x-Grade Slope	0.076		0.074	

Note: Standard errors in parentheses; + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$. The baseline model includes a cohort variable centered at 2008 (not shown), and a grade variable centered at 7.5. The outcome measure is a standardized composite of self-efficacy questionnaire items. Column 1 shows results for the full sample, while Column 2 shows results for a restricted sample of students who are observed in two or more years in the dataset.

Table A2. Trends in Student Self-Efficacy by Demographic Subgroups

	Race/Ethnicity								Socioeconomic Status			
	White		Black		Hispanic		Asian		FRPL		Non-FRPL	
Average Male Self-Efficacy												
Grade 3	0.321	***	0.246	***	0.013		0.191	***	0.046	**	0.150	***
	(0.030)		(0.030)		(0.017)		(0.031)		(0.016)		(0.023)	
Grade 4	0.389	***	0.290	***	0.067	***	0.251	***	0.098	***	0.235	***
	(0.026)		(0.024)		(0.014)		(0.023)		(0.013)		(0.020)	
Grade 5	0.406	***	0.313	***	0.113	***	0.302	***	0.138	***	0.301	***
	(0.023)		(0.022)		(0.012)		(0.022)		(0.012)		(0.018)	
Grade 6	0.379	***	0.249	***	0.073	***	0.265	***	0.104	***	0.248	***
	(0.022)		(0.020)		(0.011)		(0.019)		(0.011)		(0.018)	
Grade 7	0.231	***	0.154	***	-0.028	*	0.182	***	0.007		0.130	***
	(0.022)		(0.019)		(0.011)		(0.017)		(0.011)		(0.018)	
Grade 8	0.136	***	0.098	***	-0.069	***	0.145	***	-0.033	**	0.073	***
	(0.021)		(0.017)		(0.011)		(0.016)		(0.011)		(0.017)	
Grade 9	0.095	***	0.033		-0.074	***	0.092	***	-0.041	*	0.014	
	(0.026)		(0.022)		(0.015)		(0.022)		(0.017)		(0.019)	
Grade 10	0.003		-0.035	+	-0.091	***	0.028		-0.067	***	-0.050	**
	(0.026)		(0.021)		(0.016)		(0.024)		(0.018)		(0.018)	
Grade 11	-0.070	*	-0.032		-0.098	***	-0.065	**	-0.085	***	-0.085	***
	(0.028)		(0.021)		(0.018)		(0.022)		(0.021)		(0.018)	
Grade 12	0.071	*	-0.003		-0.016		-0.058	**	-0.012		-0.007	
	(0.030)		(0.022)		(0.021)		(0.022)		(0.024)		(0.019)	
Female-Male Self-Efficacy Gap												
Grade 3	0.109	***	0.149	***	0.165	***	0.130	***	0.163	***	0.130	***
	(0.024)		(0.028)		(0.008)		(0.027)		(0.007)		(0.019)	
Grade 4	0.049	**	0.024		0.128	***	0.090	***	0.118	***	0.063	***
	(0.015)		(0.018)		(0.006)		(0.016)		(0.005)		(0.012)	
Grade 5	-0.015		-0.013		0.051	***	-0.024		0.043	***	-0.030	*
	(0.016)		(0.019)		(0.006)		(0.017)		(0.006)		(0.013)	

Grade 6	-0.073	***	-0.078	***	-0.036	***	-0.068	***	-0.042	***	-0.057	***
	(0.016)		(0.020)		(0.007)		(0.014)		(0.006)		(0.013)	
Grade 7	-0.134	***	-0.222	***	-0.178	***	-0.123	***	-0.172	***	-0.154	***
	(0.017)		(0.020)		(0.007)		(0.016)		(0.007)		(0.014)	
Grade 8	-0.186	***	-0.284	***	-0.235	***	-0.168	***	-0.227	***	-0.213	***
	(0.018)		(0.018)		(0.007)		(0.015)		(0.006)		(0.016)	
Grade 9	-0.203	***	-0.263	***	-0.233	***	-0.172	***	-0.229	***	-0.216	***
	(0.020)		(0.020)		(0.008)		(0.015)		(0.008)		(0.015)	
Grade 10	-0.256	***	-0.250	***	-0.256	***	-0.228	***	-0.252	***	-0.255	***
	(0.022)		(0.022)		(0.009)		(0.020)		(0.008)		(0.014)	
Grade 11	-0.229	***	-0.299	***	-0.272	***	-0.222	***	-0.264	***	-0.246	***
	(0.020)		(0.021)		(0.009)		(0.018)		(0.009)		(0.013)	
Grade 12	-0.269	***	-0.245	***	-0.249	***	-0.205	***	-0.244	***	-0.232	***
	(0.021)		(0.020)		(0.010)		(0.018)		(0.009)		(0.014)	
Number of Observations	101605		103308		976308		135421		1142013		201702	
Number of Students	60147		66818		572186		80484		667894		128687	
Number of Schools	798		805		813		747		813		813	
Between-Student Intercept Variance	0.436		0.407		0.386		0.401		0.393		0.435	
Between-School Intercept Variance	0.029		0.016		0.029		0.033		0.033		0.030	
Between-School Grade Slope Variance	0.001		0.004		0.003		0.004		0.002		0.002	
Between-School Female Slope Variance	0.001		0.000		0.002		0.001		0.003		0.001	
Between-School Female-x-Grade Slope Variance	0.000		0.000		0.000		0.000		0.000		0.000	
Reliability of Student Intercept	0.586		0.488		0.512		0.593		0.519		0.580	
Reliability of School Intercept	0.202		0.153		0.556		0.283		0.617		0.276	
Reliability of School Female Slope	0.016		0.036		0.133		0.056		0.140		0.035	
Reliability of School Grade Slope	0.092		0.022		0.470		0.128		0.538		0.133	
Reliability of School Female-x-Grade Slope	0.005		0.007		0.057		0.013		0.050		0.021	

Note: The baseline model includes a fixed cohort term centered at 2008 (not shown), and a random linear grade variable centered at 7.5. The outcome measure is a composite of standardized self-efficacy questionnaire items. Standard errors in parentheses; + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Table A3. Baseline Model Results Including Student- and School-Level Variables

	Model 1				Model 2				Model 3			
	Average Male Self-Efficacy		Female-Male Self-Efficacy Gap		Average Male Self-Efficacy		Female-Male Self-Efficacy Gap		Average Male Self-Efficacy		Female-Male Self-Efficacy Gap	
Grade 3	0.335	***	0.101	***	0.687	***	0.067		0.716	***	0.08	+
	(0.022)		(0.015)		(0.071)		(0.044)		(0.070)		(0.043)	
Grade 4	0.385	***	0.07	***	0.677	***	0.024		0.7	***	0.032	
	(0.019)		(0.012)		(0.059)		(0.036)		(0.058)		(0.035)	
Grade 5	0.423	***	0.007		0.653	***	-0.051	+	0.671	***	-0.049	
	(0.016)		(0.011)		(0.054)		(0.031)		(0.052)		(0.030)	
Grade 6	0.382	***	-0.055	***	0.553	***	-0.123	***	0.565	***	-0.128	***
	(0.016)		(0.010)		(0.058)		(0.031)		(0.056)		(0.029)	
Grade 7	0.276	***	-0.164	***	0.39	***	-0.241	***	0.393	***	-0.256	***
	(0.016)		(0.010)		(0.069)		(0.035)		(0.067)		(0.032)	
Grade 8	0.228	***	-0.205	***	0.286	***	-0.291	***	0.279	***	-0.315	***
	(0.017)		(0.010)		(0.085)		(0.042)		(0.082)		(0.039)	
Grade 9	0.204	***	-0.192	***	0.202	+	-0.278	***	0.186	+	-0.309	***
	(0.021)		(0.011)		(0.104)		(0.051)		(0.100)		(0.048)	
Grade 10	0.166	***	-0.203	***	0.111		-0.297	***	0.085		-0.337	***
	(0.025)		(0.013)		(0.124)		(0.061)		(0.119)		(0.058)	
Grade 11	0.139	***	-0.197	***	0.03		-0.3	***	-0.007		-0.349	***
	(0.029)		(0.014)		(0.145)		(0.072)		(0.139)		(0.068)	
Grade 12	0.207	***	-0.164	***	0.043		-0.274	**	-0.004		-0.332	***
	(0.032)		(0.016)		(0.167)		(0.083)		(0.159)		(0.079)	
Black	-0.059	***	-0.055	***					-0.056	***	-0.06	***
	(0.011)		(0.011)						(0.008)		(0.011)	
Hispanic	-0.187	***	-0.017	*					-0.179	***	-0.029	***
	(0.010)		(0.007)						(0.006)		(0.009)	
Asian	-0.082	***	0.007						-0.079	***	0.007	
	(0.012)		(0.009)						(0.007)		(0.010)	
Other Race	-0.071	***	-0.004						-0.069	***	-0.006	
	(0.013)		(0.015)						(0.011)		(0.016)	
FRPL	-0.103	***	0.006						-0.097	***	-0.004	
	(0.007)		(0.006)						(0.004)		(0.006)	

Black-x-Grade Slope	-0.005		-0.003			-0.006	*	-0.001	
	(0.003)		(0.004)			(0.003)		(0.004)	
Hispanic-x-Grade Slope	-0.008	**	-0.01	***		-0.011	***	-0.006	+
	(0.003)		(0.003)			(0.002)		(0.003)	
Asian-x-Grade Slope	-0.003		0.004			-0.005	+	0.004	
	(0.004)		(0.003)			(0.003)		(0.004)	
Other race-x-Grade Slope	-0.01	*	0.008			-0.01	*	0.007	
	(0.004)		(0.005)			(0.004)		(0.006)	
FRPL-x-Grade Slope	0.015	***	-0.008	***		0.013	***	-0.005	*
	(0.002)		(0.002)			(0.002)		(0.002)	
Controls for Student-level Covariates	Yes				No	Yes			
Controls for School-level Covariates	No				Yes	Yes			
Number of Observations	1343715				1343715	1343715			
Number of Students	796581				796581	796581			
Number of Schools	813				813	813			
Between-Student Intercept Variance	0.395				0.400	0.395			
Between-School Intercept Variance	0.035				0.032	0.032			
Between-School Grade Slope Variance	0.002				0.002	0.002			
Between-School Female Slope Variance	0.003				0.003	0.003			
Between-School Female-x-Grade Slope Variance	0.000				0.000	0.000			
Between-School Intercept R^2	0.130				0.188	0.196			
Between-School Grade Slope R^2	0.028				0.224	0.187			
Between-School Female Slope R^2	0.041				0.061	0.067			
Between-School Female-x-Grade Slope R^2	0.080				0.213	0.148			

Note: The baseline model includes a cohort variable centered at 2008 (not shown), and a grade variable centered at 7.5. School-level variables (not shown in the table) are continuous variables for percentage of students by race/ethnicity and socioeconomic disadvantage, averaged within schools across survey years (2014-15 through 2016-17). Reference categories are White, non-FRPL students for Models 1 and 3. Standard errors in parentheses; + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.

Table A4. Summary of Student and School Covariate Data for Multi-Year Test Score Sample

	Mean	SD
Panel 1. Student-level Summary Statistics		
Female	0.50	(0.50)
Male	0.50	(0.50)
White	0.08	(0.26)
Black	0.07	(0.25)
Hispanic	0.74	(0.44)
Asian	0.09	(0.29)
Other	0.02	(0.14)
FRPL	0.87	(0.33)
Self-Efficacy Composite	3.62	(1.00)
Number of Students	246174	
Panel 2. School-level Summary Statistics		
White (%)	0.09	(0.13)
Black (%)	0.08	(0.11)
Hispanic (%)	0.72	(0.25)
Asian (%)	0.09	(0.14)
Other (%)	0.02	(0.03)
FRPL (%)	0.80	(0.20)
High-Scoring (%)	0.25	(0.16)
Low-Scoring (%)	0.25	(0.12)
Mid-Scoring (%)	0.51	(0.08)
Average Enrollment per Grade (100s)	1.47	(1.16)
Culture & Climate Survey Constructs		
Belonging Composite	3.50	(0.21)
Learning Composite	3.63	(0.24)
Discipline Composite	3.49	(0.22)
Number of Schools	666	

Note: The self-efficacy composite is an average of four items that pertain to this construct. The culture & climate survey constructs are composites that measure school averages of standardized items pertaining to each construct. The multi-year test score sample is restricted to 4th- 8th grade students with nonmissing math and self-efficacy scores over two consecutive years at a time (e.g., 2014–15 and 2015–16)

Table A5. Trends in Student Self-Efficacy by Year and Achievement Subgroup

	High-Scoring		Low-Scoring		Mid-Scoring	
Average Male Self-Efficacy						
Grade 4	0.530	***	-0.099	***	0.148	***
	(0.011)		(0.013)		(0.010)	
Grade 5	0.570	***	-0.100	***	0.173	***
	(0.010)		(0.012)		(0.009)	
Grade 6	0.561	***	-0.129	***	0.132	***
	(0.012)		(0.013)		(0.010)	
Grade 7	0.486	***	-0.210	***	0.030	**
	(0.013)		(0.014)		(0.011)	
Grade 8	0.461	***	-0.287	***	-0.018	
	(0.013)		(0.014)		(0.011)	
Female-Male Self-Efficacy Gap						
Grade 4	0.087	***	0.110	***	0.107	***
	(0.013)		(0.015)		(0.010)	
Grade 5	0.008		0.044	**	0.028	**
	(0.011)		(0.013)		(0.009)	
Grade 6	-0.074	***	-0.053	***	-0.068	***
	(0.012)		(0.015)		(0.010)	
Grade 7	-0.173	***	-0.229	***	-0.216	***
	(0.012)		(0.014)		(0.009)	
Grade 8	-0.189	***	-0.267	***	-0.249	***
	(0.012)		(0.015)		(0.010)	
Number of Students	99143		99221		200386	
Number of School-Year Observations	1899		1892		1923	
Number of Schools	665		663		666	

Notes: Standard errors in parentheses. + p <0.10 * p < 0.05 ** p < 0.01 *** p < 0.001

Table A6. Associations between Grade-to-Grade Changes in Self-Efficacy and Test Scores

	Change in Test Scores					
	All Students		Females		Males	
3rd to 4th Grade						
Change in Efficacy	0.015	***	0.006		0.016	***
	(0.004)		(0.004)		(0.004)	
Female-x-Change in Efficacy	-0.010	+				
	(0.005)					
Number of Students	71170		35509		35661	
Number of Schools	398		380		388	
4th to 5th Grade						
Change in Efficacy	0.019	***	0.010	**	0.019	***
	(0.004)		(0.004)		(0.004)	
Female-x-Change in Efficacy	-0.008					
	(0.005)					
Number of Students	93942		46794		47148	
Number of Schools	527		527		526	
5th to 6th Grade						
Change in Efficacy	0.018	**	0.023	***	0.018	**
	(0.005)		(0.005)		(0.006)	
Female-x-Change in Efficacy	0.004					
	(0.008)					
Number of Students	76016		38300		37716	
Number of Schools	299		299		299	
6th to 7th Grade						
Change in Efficacy	0.049	***	0.046	***	0.049	***
	(0.005)		(0.005)		(0.005)	
Female-x-Change in Efficacy	-0.002					
	(0.007)					
Number of Students	80659		40781		39878	
Number of Schools	179		179		179	
7th to 8th Grade						
Change in Efficacy	0.062	***	0.042	***	0.062	***
	(0.006)		(0.006)		(0.006)	
Female-x-Change in Efficacy	-0.021	*				
	(0.008)					
Number of Students	76963		38512		38451	
Number of Schools	180		180		180	

Note: The grade-to-grade changes in self-efficacy have been standardized to a N(0,1) distribution; the grade-to-grade changes in test scores have been standardized to a N(0,1) distribution. Selected model coefficients shown. Standard errors in parentheses; + $p < 0.10$ * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$.